Environment Overview

1. **Describe the deterministic and stochastic environments, which were defined (set of actions/states/rewards, main objective, etc).**

2. **Provide visualizations of your environments.**

The environment represents a grid world game called GoldHunter. In this game a hunter starts the game from the cell (0, 0) and should try to maximize his score after finishing the game. The game generally consists of multiple objects which each of them have different properties and behaviours. An overview of the map of the game is shown in the below image:

First, I start by defining the environment by its fundamental components and then we'll dive into the dynamics of the game.

States = {49 positions on the grid world}
Actions = {UP, RIGHT, DOWN, LEFT, LOG_UP, LOG_RIGHT, LOG_DOWN, LOG_LEFT}
Rewards = { -5, -3, -1, +5, +10}

Objects

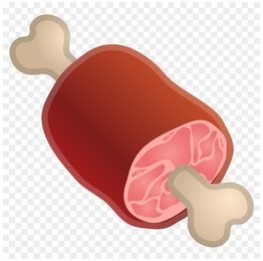
Hunter

Hunter is our agent and is trying to reach the gold, while maintaining his power. He can maintain his power **by avoiding traps, reaching to gold faster or eating food**

Gold

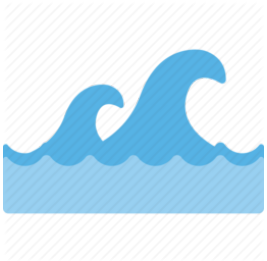
Obviously this is our goal. By reaching the cell containing gold the episode terminates and agent receive a reward of 10

Food



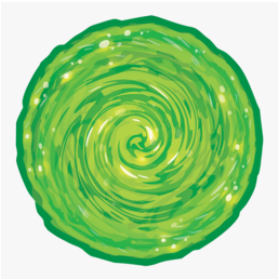
Food is beneficial for the hunter because it keeps him from getting hungry :(
Although, hunter can find the gold and finish the game, but, collecting food before that maximizes his reward.

Stream



Streams can be both dangerous and beneficial! When the agent falls into one, the stream takes him to the last place it goes. If the stream ends at the border of grid world the agent falls on the left or right of the stream!

Teleport



Teleport is a non deterministic element of this world. It connects the lower right to the higher left; however, it is non deterministic and it might not get activated when the hunter first goes inside it.

Trap

Traps are tricky creatures! When the hunter falls into one, he gets hurt and gets a reward of -3. But that's not the whole story! The agent can run out of traps only when he performs actions in opposite directions (e.g. right and left) then he will be free.

Log

Logs are objects that the hunter can craft around him. Hunter has 4 other actions other than his movement actions to each direction. These actions craft a log around him in each direction and lets him pass the object ahead of him without being affected by its properties. For example, he can build a log on the stream or a trap and pass on it. However, **crafting is costly and makes him tired**, so he'll receive a reward of -3 when he builds the log.

Dynamics of the environment

3. How did you define the stochastic environment?

Deterministic environment

Agent starts at (0, 0) and tries to get the gold by going to (6, 6). In this journey, the agent might fall into traps which he can run out of by applying movements to opposite directions or he can craft a log on top of the trap, before falling into it.
Also there is the possibility of falling into the stream which as described above, takes him to the end of stream. He can get food and improve his reward and finally reach the gold.

Stochastic environment

To add stochasticity to the environment, I added a teleport channel, which can help the agent teleport to the other corner of the map where he can access food! The teleport lets the hunter move to the other corner with probability of P.

4. What is the difference between the deterministic and stochastic environments?

In deterministic environment, given that you are in state S and you perform action a you will end up in a specific state S' and receive reward r. Although, In stochastic environments, there is a probability for this transaction and you might end up in a different state and receive a different reward as well.

In GoldHunt environment as an example the following stochastic behaviors exist:

$$P([0,6],\ +5 \mid [5,0],\ RIGHT) = P$$
$$P([6,0],\ -1 \mid [5,0],\ RIGHT) = 1 - P$$
$$P([0,6],\ +5 \mid [6,1], DOWN) = P$$
$$P([6,0],\ -1 \mid [5,0],\ DOWN) = 1 - P$$